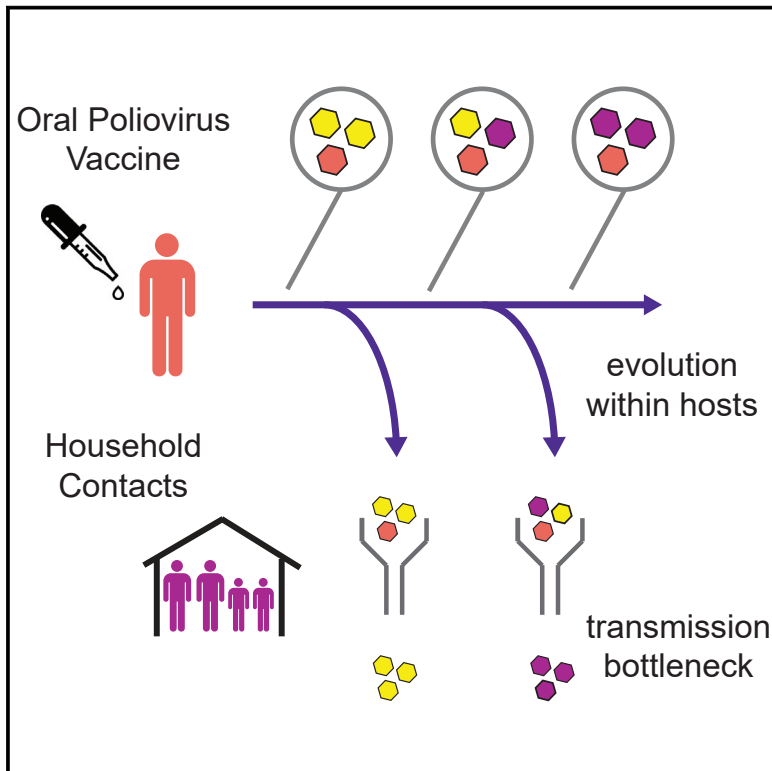


# Cell Host & Microbe

## The Early Evolution of Oral Poliovirus Vaccine Is Shaped by Strong Positive Selection and Tight Transmission Bottlenecks

### Graphical Abstract



### Authors

Andrew L. Valesano, Mami Taniuchi, William J. Fitzsimmons, ..., Wesley Wong, Michael Famulare, Adam S. Lauring

### Correspondence

alauring@med.umich.edu

### In Brief

The emergence of vaccine-derived polioviruses through evolution of the oral polio vaccine poses a significant obstacle to global poliovirus eradication. Valesano et al. use sequencing of samples from vaccine recipients and household contacts to identify multiple mutations that are selected within hosts. Tight bottlenecks limit the onward transmission of these variants.

### Highlights

- We used deep sequencing to define the evolutionary trajectories of type 2 OPV
- We identified strong positive selection at multiple, non-attenuating sites
- Transmission bottleneck limits spread of variants that are selected within hosts
- This work provides insights into the evolution of live attenuated virus vaccines



Article

# The Early Evolution of Oral Poliovirus Vaccine Is Shaped by Strong Positive Selection and Tight Transmission Bottlenecks

Andrew L. Valesano,<sup>1,2</sup> Mami Taniuchi,<sup>3</sup> William J. Fitzsimmons,<sup>4</sup> Md Ohedul Islam,<sup>5</sup> Tahmina Ahmed,<sup>5</sup> Khalequ Zaman,<sup>5</sup> Rashidul Haque,<sup>5</sup> Wesley Wong,<sup>6</sup> Michael Famulare,<sup>6</sup> and Adam S. Lauring<sup>4,7,8,\*</sup>

<sup>1</sup>Program in Cellular and Molecular Biology, University of Michigan, Ann Arbor, MI 48109, USA

<sup>2</sup>Medical Scientist Training Program, University of Michigan, Ann Arbor, MI 48109, USA

<sup>3</sup>Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA 22908, USA

<sup>4</sup>Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA

<sup>5</sup>International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh

<sup>6</sup>Institute for Disease Modeling, Global Good, Bellevue, WA 98005, USA

<sup>7</sup>Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA

<sup>8</sup>Lead Contact

\*Correspondence: [alauring@med.umich.edu](mailto:alauring@med.umich.edu)

<https://doi.org/10.1016/j.chom.2020.10.011>

## SUMMARY

The emergence of circulating vaccine-derived polioviruses through evolution of the oral polio vaccine (OPV) poses a significant obstacle to polio eradication. Understanding the early genetic changes that occur as OPV evolves and transmits is important for preventing future outbreaks. Here, we use deep sequencing to define the evolutionary trajectories of type 2 OPV in a vaccine trial. By sequencing 497 longitudinal stool samples from 271 OPV2 recipients and household contacts, we were able to examine the extent of convergent evolution in vaccinated individuals and the amount of viral diversity that is transmitted. In addition to rapid reversion of key attenuating mutations, we identify strong selection at 19 sites across the genome. We find that a tight transmission bottleneck limits the onward transmission of these early adaptive mutations. Our results highlight the distinct evolutionary dynamics of live attenuated virus vaccines and have important implications for the success of next-generation OPV.

## INTRODUCTION

Genetic reversion and the associated loss of attenuation in oral poliovirus vaccine (OPV) strains are major barriers to achieving global poliovirus eradication (Kew et al., 2005). In areas of low vaccine coverage, OPV can evolve into circulating vaccine-derived polioviruses (cVDPV) that cause cases of poliomyelitis that are indistinguishable from those caused by wild polioviruses (WPV) (Jenkins et al., 2010; Kew et al., 2002; Pons-Salort et al., 2016). Of the three OPV serotypes, the Sabin type 2 is responsible for most cVDPV outbreaks (Burns et al., 2014; Kew and Palansch, 2018). Following the eradication of wild-type 2 polioviruses, the Global Polio Eradication Initiative switched routine immunization schedules from trivalent OPV (tOPV; Sabin types 1, 2, and 3) to bivalent OPV (bOPV; Sabin types 1 and 3) to reduce the risk of future cVDPV2 outbreaks. However, monovalent type 2 OPV (mOPV2) is still used to combat cVDPV2 outbreaks. While the global replacement of tOPV with bOPV has reduced the presence of OPV2 in surveillance samples (Blake et al., 2018), cVDPV2 outbreaks remain a major problem. This is partly due to the continued reliance on monovalent Sabin

type 2 OPV (mOPV2) to control new cVDPV2 outbreaks. Nearly half of the cVDPV2 outbreaks observed after the withdrawal of tOPV resulted from a previous mOPV2 intervention response (Macklin et al., 2020).

The emergence of cVDPV is a recurrent evolutionary process that exhibits a high degree of parallel, or convergent, evolution. Most data on cVDPV come from poliovirus isolates in cases of acute flaccid paralysis or environmental surveillance (Burns et al., 2013; Famulare et al., 2016; Shaw et al., 2018; Stern et al., 2017). A recent study of type 2 cVDPV sequences from multiple outbreaks in five countries identified a limited number of sites under positive selection across independent lineages (Stern et al., 2017). Three mutations—A481G, U2909C (VP1-I143T), and U398C—were inferred to be under the strongest selection pressure and precede subsequent substitutions. The A481G and U398C mutations are located in the 5' noncoding region and are functionally important to RNA structures in the internal ribosome entry site (IRES). All three mutations are known molecular determinants of attenuation, occur within the first 2 months after vaccination, and are associated with increased virulence in animal models (Famulare et al., 2016; Macadam



et al., 1991, 1993; Muzychenko et al., 1991; Ren et al., 1991; Stern et al., 2017). For these reasons, they are referred to as “gatekeeper” mutations that initiate the process of attenuation loss. Although phylogenetic studies have provided important information on the evolutionary trajectories of cVDPV, they are limited in their ability to resolve the exact timing of gatekeeper mutations and may lack power to detect natural selection due to sampling bias (Geoghegan and Holmes, 2018). Isolates of cVDPV have undergone months or years of evolution prior to isolation and lack a definitive link to the time of vaccine administration, further limiting our understanding of the early evolution of OPV in humans.

Investigating the evolutionary dynamics within individual hosts can complement phylogenetic studies of virus evolution (Lauring, 2020). Complex evolutionary processes that take place within the span of a single infection cannot be resolved by standard consensus sequencing. Individual mutations, most often single-nucleotide variants, arise within infected hosts and change in frequency according to the forces of natural selection and genetic drift. Studying how viruses evolve at this scale can uncover genomic sites under selective pressure, clarify the relative roles of selection and drift in viral evolution, and can inform sequence-based diagnostic and surveillance tools (Dolan et al., 2018b; Holubar et al., 2019).

Various approaches have been used to study the molecular epidemiology of poliovirus and to monitor OPV stocks for reversion (Neverov and Chumakov, 2010; Sarcey et al., 2017), but few have been purposed for measuring viral diversity within naturally infected hosts. Routine surveillance for VDPV involves sequencing only the region encoding the capsid protein VP1 (Kilpatrick et al., 2011). Many approaches for whole genome sequencing rely on amplification of viral isolates in cell culture, which may not accurately preserve the diversity present in the original specimen (Montmayeur et al., 2017). Other high-throughput sequencing approaches that aim to measure within-host diversity have targeted only a specific portion of the poliovirus genome (Sahoo et al., 2017). While some have sequenced viral genomes or specific genomic regions from asymptomatic vaccine recipients (Boot et al., 2007; Dedepsidis et al., 2006; van der Sanden et al., 2009; Stern et al., 2017), we lack a comprehensive characterization of the early evolutionary dynamics of OPV within vaccinated individuals and during transmission to their close contacts.

Here, we used whole genome, deep sequencing of stool samples from a clinical trial of OPV to elucidate the early evolution of polioviruses within and between human hosts. We developed an approach for sequencing OPV genomes directly from primary stool samples and validated its accuracy for identification of intrahost single-nucleotide variants (iSNV). We applied this approach to samples from a recent trial that investigated the effect of tOPV cessation on the transmission of type 2 OPV (Taniuchi et al., 2017). The trial included a defined point of introduction of monovalent type 2 OPV (mOPV2) and weekly longitudinal sampling of vaccine recipients and their household contacts; it therefore represents an opportunity to investigate the early evolutionary dynamics of OPV2 in a community setting. We identified several mutations under strong positive selection, most of which are located in the capsid proteins and the 5′ noncoding region. By comparing viral diversity across household transmis-

sion pairs, we found that mOPV2 experiences a narrow transmission bottleneck, which may limit the spread of mutations that are strongly selected within hosts. These results connect the within-host selection of mutations with the dynamics of viral transmission and enhance our understanding of cVDPV evolution.

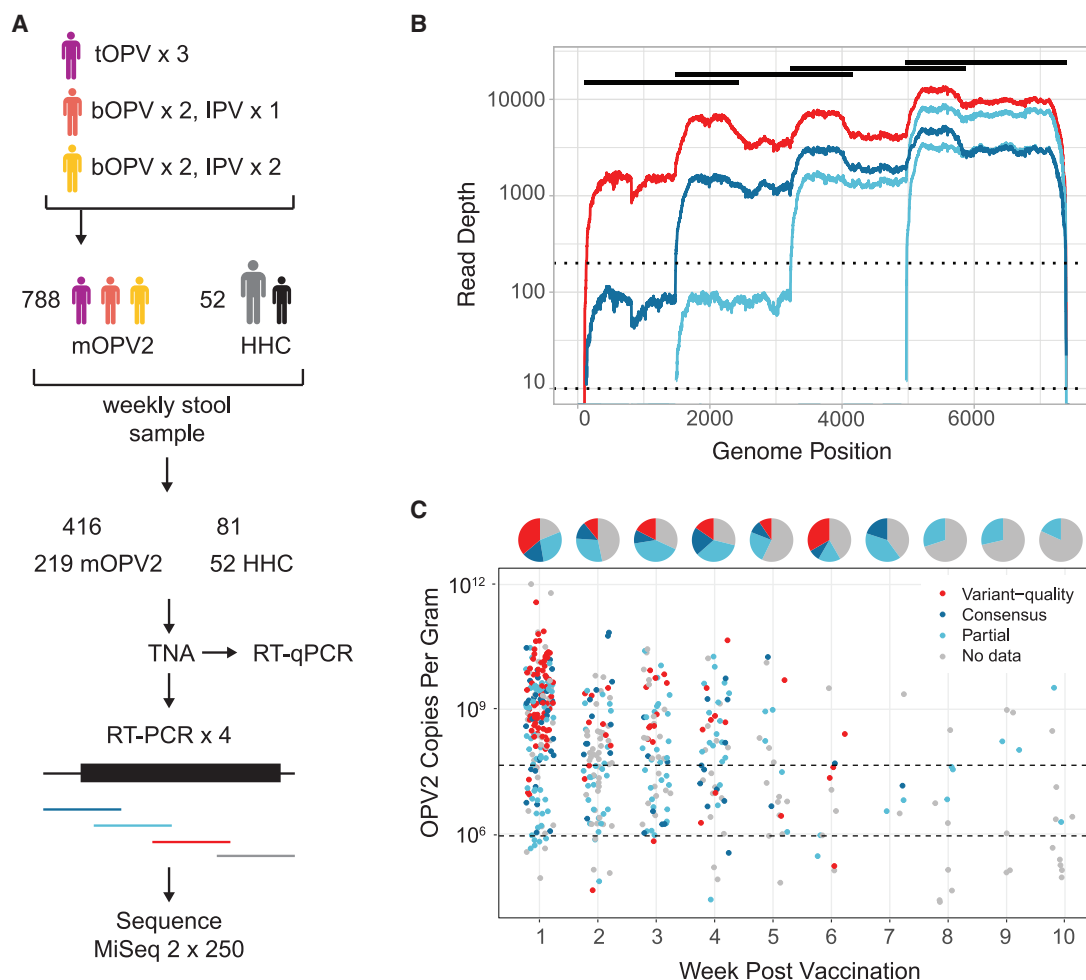
## RESULTS

We used high-depth-of-coverage sequencing on the Illumina platform to define the within-host diversity of mOPV2 in samples from vaccinated individuals and their household contacts (Taniuchi et al., 2017). These samples were collected as part of a cluster-randomized trial of OPV in the rural Matlab region, where the International Center for Diarrheal Disease Research, Bangladesh (icddr,b) has conducted demographic and public health research since the 1960s (Alam et al., 2017). The trial assessed the impact of tOPV withdrawal on OPV community transmission by randomizing 67 villages to three different vaccination schedules: tOPV, bOPV followed by one dose of inactivated polio vaccine (IPV), and bOPV followed by two doses of IPV. The trial then implemented a coordinated mOPV2 vaccination campaign over the course of 1 week, targeting 40% of children under 5 years of age. Shedding of OPV types 1–3 from 800 individuals across the three arms was monitored by quantitative RT-PCR of weekly stool samples. Transmission was measured by monitoring stool samples in household contacts of mOPV2 recipients. We selected 497 specimens from the vaccination campaign period for genome amplification and sequencing, prioritizing those with a stool viral load  $>10^6$  copies/gram.

### Sample Sequencing and Assessment of Genome Coverage

We sequenced 416 samples from 219 mOPV2 recipients and 81 samples from 52 household contacts (Figure 1A). We amplified poliovirus genomes as overlapping RT-PCR amplicons using degenerate primers that recognize all three OPV serotypes (Table S1). We performed separate RT-PCR reactions for each segment and pooled them prior to library preparation (see Materials and Methods). Given that a low viral titer influences the accuracy of within-host variant identification (McCrone and Lauring, 2016), we amplified and sequenced samples with an OPV2 viral load between  $9 \times 10^5$  copies/gram and  $4.5 \times 10^7$  copies/gram in duplicate. These cutoffs are based on the distribution of viral loads across the cohort and our empirically defined viral load cut-offs for influenza virus (McCrone et al., 2018). We amplified and sequenced several samples below  $9 \times 10^5$  copies/gram that were collected from household contacts.

Depth of coverage across the OPV2 genome for a given sample was uneven (Figure 1B). The 3′ end generally exhibited coverage of at least an order of magnitude greater than the 5′ end, which contains the highly structured IRES (Lévêque and Semler, 2015). Three hundred twenty-seven samples had greater than  $10\times$  coverage of at least one of the four amplicons (Figure 1C, light blue points), and 179 samples had greater than  $10\times$  coverage across the whole genome (Figure 1C, dark blue points). We identified 111 samples with  $>200\times$  coverage (Figure 1C, red points), 81 samples with  $>500\times$  coverage, and 48 samples with  $>1000\times$  coverage across the genome, based on averages across a 50-bp sliding window. The majority of



**Figure 1. Overview of Study and Sequence Data**

(A) Schematic of study design and sample processing. The clinical trial had three arms with lead-up vaccination as indicated. tOPV, trivalent OPV; bOPV, bivalent OPV; IPV, inactivated polio vaccine. All individuals (n = 788) then received mOPV2. Stool samples were collected weekly from mOPV2 recipients and household contacts (HHC). Only 52 household contacts had detectable shedding of OPV2. Total nucleic acid (TNA) was extracted from stools. Poliovirus genomes were amplified from each sample as four overlapping RT-PCR amplicons. For each sample, these amplicons were pooled and prepared for sequencing.

(B) Line graph of sequencing coverage of four selected samples in three coverage groups. Log<sub>10</sub> of coverage depth on the y axis and genome position on the x axis. One variant-quality sample shown in red, one consensus-quality sample shown in dark blue, and two partial-genome samples shown in light blue. Amplicons are shown as black bars (top). Dotted lines show cutoffs at 200x and 10x used for defining coverage groups.

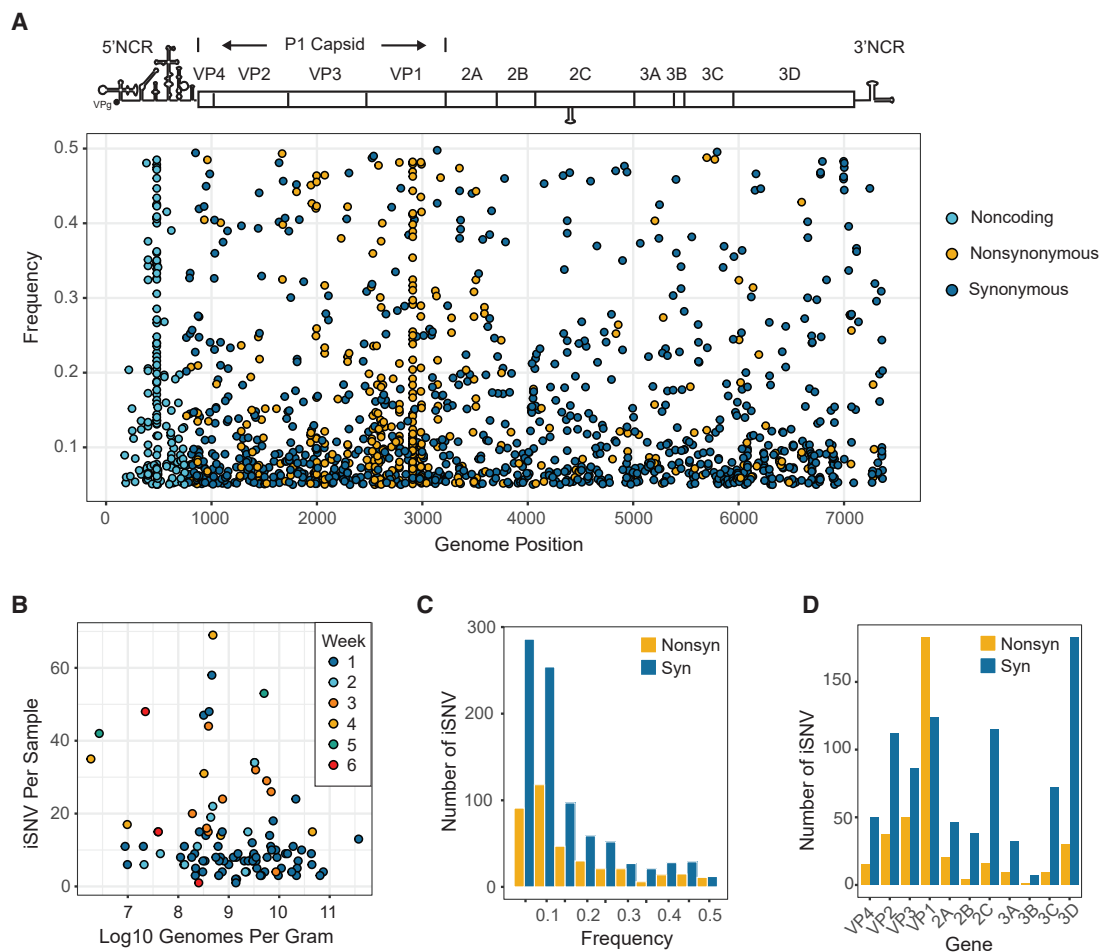
(C) Coverage groups of samples sequenced in this study. Each sample is shown as a point with OPV2 copies per gram of stool on the y axis and weeks post-vaccination on the x axis. Pie charts above each week indicate the proportion of samples with variant-quality data (red), consensus quality data (dark blue), partial genome sequence data (light blue), and no data (gray). The region in between the dotted lines shows the samples that were sequenced in duplicate.

samples that yielded at least partial OPV2 genome coverage were collected in the first 2 months following vaccination (Figure 1C), which is consistent with the known shedding duration of Sabin type 2 (Famulare et al., 2018). Most individuals were represented by only one sample, although a subset of individuals had multiple longitudinal samples with at least partial genome data (Figure S1).

### Empiric Evaluation of Variant Calling Criteria

We benchmarked the accuracy of our variant calling criteria for iSNV identification by sequencing defined mixtures of WPV1 (Mahoney strain) and OPV1 in stool-derived total nucleic acid (TNA) with viral genome concentrations ranging from  $9 \times 10^4$  copies/gram to  $4.5 \times 10^7$  copies/gram of stool. These concen-

trations were tailored to match those of the sequenced clinical samples. We then calculated the sensitivity and specificity of iSNV identification at various thresholds of input concentration, sequencing coverage, and iSNV frequency (Table S2). At a coverage depth of 200x, we reliably identified the expected single-nucleotide variants at 5% frequency with 95% sensitivity at all genome-copy inputs. However, at this coverage level, sensitivity was weaker for low frequency variants. We found that the number of false positives was low when the viral load was greater than  $4.5 \times 10^7$  genome copies/gram. Specificity declined at viral loads below this cutoff, with a false positive rate of ~1% at all coverage levels. While some of these false positives can be filtered with various criteria, such as base and mapping quality, performing technical replicates proved to be the most effective



**Figure 2. Within-Host Diversity in 101 Variant-Quality Samples from mOPV2 Vaccine Recipients**

(A) Minor iSNV shown as points, with frequency on the y axis and genome position on the x axis. Non-coding iSNV are shown in light blue, nonsynonymous iSNV in yellow, and synonymous iSNV in dark blue.

(B) Number of minor iSNV (y axis) versus  $\log_{10}$  of genome copies per gram of stool (x axis). Color of each point is shown by the week post-vaccination of sample collection.

(C) Histogram of minor iSNV in polyprotein by frequency with bin width of 0.05. Nonsynonymous iSNV are shown in yellow and synonymous iSNV in dark blue.

(D) Histogram of minor iSNV by protein-coding region in the polyprotein. Nonsynonymous iSNV are shown in yellow and synonymous iSNV in dark blue.

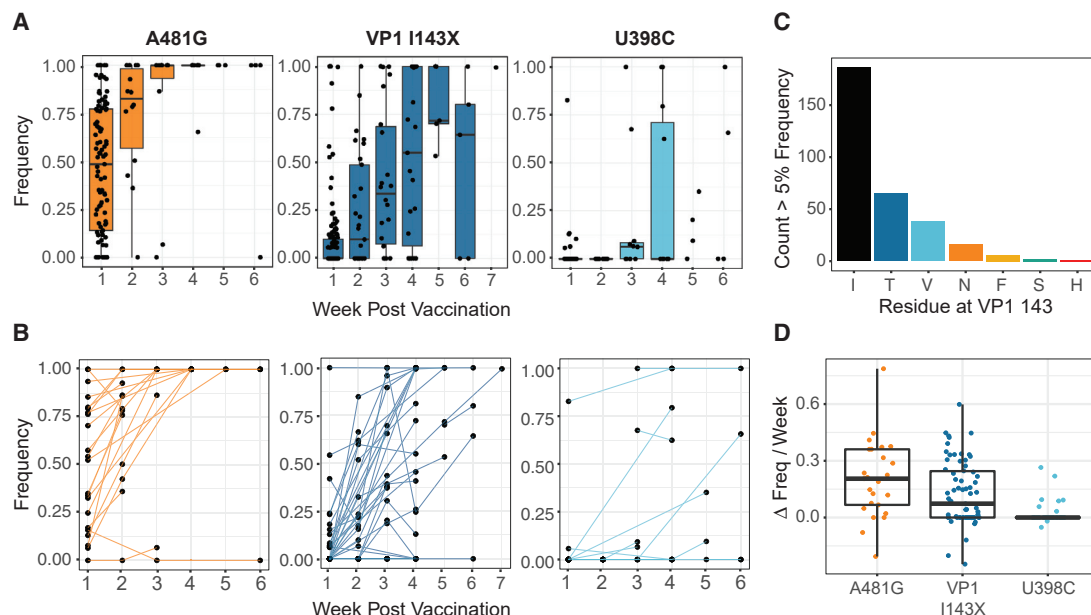
approach for removing false positive variants. When considering only variants that were identified in both sequencing replicates, specificity dramatically improved even at low viral loads. Overall, these data validate our approach for poliovirus sequencing and demonstrate high sensitivity and specificity for variants above 5% frequency in the majority of sequenced samples. Therefore, we identified iSNV above a frequency threshold of >5% in 111 samples that had >200 $\times$  coverage by sliding window across the genome, denoted here as “variant quality.” For analyses of variants at particular genome positions, we included samples with >200 $\times$  mean coverage in a 50-bp window containing the site of interest.

### Signatures of Selection

We characterized within-host diversity in 101 variant-quality samples from mOPV2 recipients. Minor iSNV (<50% frequency) were dispersed across the genome in both the 5' noncoding region and the polyprotein, with greater numbers of variants at

lower frequencies (Figure 2A). Each sample contained a median of 9 minor iSNV (IQR 6–15). There were more minor iSNV per sample with greater time since vaccination (Figure 2B). This association remained significant even after we controlled for the time-varying factor of viral load, which can affect iSNV identification ( $p < 0.001$ , multiple linear model). Our estimates of minor iSNV frequency were consistent when compared between technical replicates of 11 variant-quality samples (adjusted  $r$ -squared = 0.763, Figure S2A). While previous work has shown that the measured frequency of a variant can be affected by mutations in the primer binding sites (Grubaugh et al., 2019), we were unable to distinguish this effect from the overall error in our frequency measurements. We identified a greater proportion of synonymous iSNV relative to nonsynonymous iSNV (ratio 0.43, Figure 2C). However, in the VP1 capsid subunit, there was an enrichment of nonsynonymous minor iSNV compared to other protein coding regions (Figure 2D). We calculated the dN/dS ratio with samples from mOPV2 recipients that had full





**Figure 3. Selection of Gatekeeper Mutations in Vaccine Recipients**

(A) Frequency of A481G, VP1-143X, and U398C by time from vaccination. Each point represents one sample, and boxplots are shown for weeks with five or more data points. Boxplots represent the median and 25<sup>th</sup> and 75<sup>th</sup> percentiles, with whiskers extending to the most extreme point within the range of the median  $\pm 1.5$ -times the interquartile range.

(B) Frequency of A481G, VP1-143X, and U398C by time from vaccination. Each point represents one sample, with lines connecting samples from the same individual.

(C) Barplot showing the number of samples with the indicated residues present at a frequency of 5% or above at VP1-143.

(D) Change in frequency per week of three gatekeeper mutations prior to reaching fixation. Boxplots represent the median and 25<sup>th</sup> and 75<sup>th</sup> percentiles, with whiskers extending to the most extreme point within the range of the median  $\pm 1.5$ -times the interquartile range.

consensus sequences across the polyprotein ( $n = 157$ ). VP1 had the highest dN/dS ratio compared to the rest of the protein coding regions (Table S3).

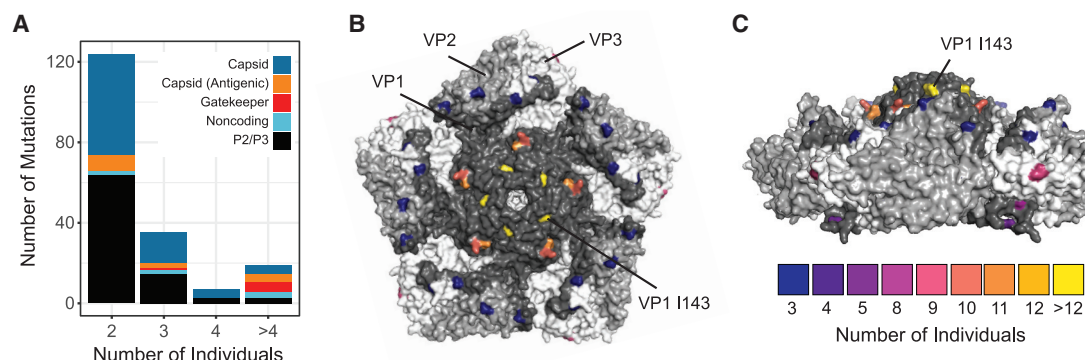
### Positive Selection of Gatekeeper Mutations

The dN/dS ratio is an imperfect metric for detecting selection, particularly within hosts, and it is unable to identify positive selection of mutations in noncoding regions (Kryazhimskiy and Plotkin, 2008). While changes in frequency of viral variants can be caused by multiple evolutionary forces, observing the same mutation arise in independent viral populations is suggestive of positive selection (Dolan et al., 2018a; Gutierrez et al., 2019). We therefore analyzed the mutational dynamics at three positions that are major attenuating sites—positions 481 and 398 in the 5' noncoding region and codon 143 of VP1 (nucleotide positions 2908–2910). We used our time-series data to directly measure the frequency changes of the gatekeeper mutations in vaccine recipients. All three were present in several individuals within the first week of vaccination. A cross-sectional analysis of mutation frequency as a function of time demonstrated fixation of A481G within 2–3 weeks and U2909C in about 5 weeks, although there was substantial interindividual variability (Figures 3A and 3B). While A481G reached consensus in 11 of 14 samples by week 2, U2909C reached consensus in only 5 of 22 samples by week 3. VP1-143 most frequently reverted from isoleucine to threonine, but several other alternative residues were present (Figure 3C). Data from individuals with more than one sequenced

sample demonstrated a rapid increase in frequency. Although mutation frequency occasionally decreased, presumably due to stochastic effects, these mutations increased in nearly all individuals (Figures 3B and 3D). We applied a beta regression model to estimate the time to fixation in the population. For mutations A481G, U2909C, and U398C, the model predicts a frequency of  $>0.5$  at weeks 2, 5, and 12, respectively, and  $>0.95$  by weeks 6, 13, and 46, respectively (Figure S3). While we had more data from individuals in the bOPV/IPV study arms, each mutation rose in frequency over the same time interval regardless of vaccination history. This suggests that the selection for these mutations is not substantially driven by the presence of mucosal immunity generated by tOPV (Figure S3). Overall, our time-series data on mOPV2 recipients show rapid fixation of mutations at key attenuating sites in the first several weeks post-vaccination.

### Additional Sites with Positive Selection

Next, we identified non-gatekeeper mutations that occurred independently across mOPV2 recipients. We restricted our analysis to 83 individuals who received mOPV2 and for whom we had at least one variant-quality sample. While most mutations were unique to a given viral population, a large number of mutations were present at  $\geq 5\%$  frequency in  $\geq 2$  individuals (Figure 4A). We performed a permutation test to quantitatively assess whether this distribution could occur due to chance alone. We drew random sites across the genome and tallied the number



**Figure 4. Mutations Arising in Multiple mOPV2 Vaccine Recipients**

(A) Stacked barplot of the number of mutations identified (y axis) by the number of individuals with each mutation (x axis). Mutations in  $\geq 3$  individuals were statistically significant by permutation test, see text. Colors show the category of each mutation.

(B and C) (B) Structure of type 2 poliovirus capsid pentamer (PDB: 1EAH) and side view (C), where highlighted residues are color-coded by number of mOPV2 vaccine recipients with nonsynonymous substitutions at that amino acid site.

of sites shared across multiple individuals (see STAR Methods). We observed more shared mutations than would be expected by chance for mutations in  $\geq 3$  individuals (Figure 4A). This result is robust to the assumption that all genome sites can be mutated; reducing the fraction of sites available for mutation did not affect statistical significance until the fraction dropped below 50% (Figure S4).

Excluding the gatekeeper mutations, we found 19 mutations that were present in  $\geq 4$  individuals (Table 1). Two mutations, G491A and G619U, were located in the IRES. Five mutations were located outside the capsid (P1) in polypeptide regions P2 and P3. The capsid proteins (VP1–VP4) were highly represented (12 of 19 mutations), with four mutations encoding nonsynonymous mutations in known antigenic sites. Mutation A2986G encodes a nonsynonymous substitution, VP1-K169E, in antigenic region NAg1. Two more mutations in NAg1, G2782A and C2783A, encode nonsynonymous changes at VP1-101 (A101T and A101D, respectively). Mutation A1997G, found in 8 of 83 individuals, encodes VP3-H77R in antigenic region NAg3b. This mutation was identified in a previous phylogenetic study as having intermediate evidence for positive selection across cVDPV lineages (Stern et al., 2017). Our host-level data show that this mutation is indeed under strong positive selection. We did not detect U2523C, C2006A, U1376A, or U3320A, which are predicted to occur  $>2$  months after vaccination (Stern et al., 2017).

We also examined independent capsid mutations at the codon level, such that distinct mutations at the same amino acid site were included. We identified 19 amino acid sites at which three or more individuals exhibited nonsynonymous substitutions at a frequency of  $>5\%$ . Many of these sites mapped to the surface of the type 2 capsid (Figures 4B and 4C). In addition to the amino acid sites specified above, this analysis revealed four more antigenic sites with parallel nonsynonymous substitutions: VP3-58 (T58I and T58A) and VP1-T291A in NAg3a, VP2-N72D in NAg3b, and VP1-S222P in NAg2. Although dN/dS analysis is often not sensitive enough to identify positive selection at specific sites over short time scales, amino acid sites VP1-143 and VP1-101 had a dN/dS ratio greater than 1 based on analysis of consensus genomes ( $\text{Pr}(\omega > 1) > 0.95$ , Bayes empirical Bayes method, Table S3).

Together, our results demonstrate rapid positive selection of mutations in the IRES and exposed sites in the OPV2 capsid.

We sought to determine whether mutations selected early in OPV2 evolution persist in genomes of neurovirulent cVDPV. We queried alignments of cVDPV from cases of acute flaccid paralysis for the presence of the mutations identified here (Table 1). As expected, the gatekeeper mutations were reliably detected in nearly all cVDPV genomes. Of the other 19 mutations, we found 16 in at least 1% of cVDPV genomes queried. We detected some mutations, including G491A, C2580U, and U1641C, in at least 10% of cVDPV genomes. While it is unknown how these mutations may contribute to the development of cVDPV outbreaks, our data show that mutations that recur in divergent cVDPV lineages can be identified very early in OPV2 evolution.

### Estimation of the Transmission Bottleneck

Transmission bottlenecks influence the rate of adaptation and are important for understanding viral evolution in host populations (Elena et al., 2001). The OPV vaccine trial included stool samples collected from the household contacts of mOPV2 recipients. Shedding in these individuals allowed us to analyze the extent of viral diversity that is transmitted to new hosts. We identified four transmission pairs for which we had sequence data from both the donor and recipient collected within 1 week of each other (Table S4); low viral loads, especially in household contacts, thwarted sequencing of more putative household pairs. In each case, transmission occurred within the first 3 weeks after vaccination, consistent with the known magnitude and duration of OPV shedding. We obtained variant-quality samples from donor and recipient of one pair, while the rest had either consensus or partial genomes with regions of variant-quality coverage.

We compared within-host diversity across donors and recipients using only the genomic regions that had a depth of coverage sufficient for identification of within-host variants in both samples. There were few polymorphic sites shared across hosts in these household pairs (Figure 5A). While major variants ( $>50\%$  frequency) in the donor were usually found in the recipient, most minor variants were not found in the recipient, suggestive of a narrow transmission bottleneck.

**Table 1. Mutations Identified in Multiple Individuals**

Mutation <sup>a</sup>	Individuals <sup>b</sup>	Group	Type <sup>c</sup>	Region	Fraction of cVDPV <sup>d</sup>
A481G	72	Gatekeeper	Noncoding	5' UTR	1
U2909C	25	Gatekeeper	NS	VP1	0.75
A2908G	19	Gatekeeper	NS	VP1	0.03
U398C	16	Gatekeeper	Noncoding	5' UTR	0.94
A2074G	12	Capsid	NS	VP3	0.01
A2992G	11	Capsid	NS	VP1	0.03
A2986G	10	Antigenic	NS	VP1	0.01
G6084U	10	3D	S	3D	0.03
A1997G	8	Antigenic	NS	VP3	0.05
U2909A	8	Gatekeeper	NS	VP1	0.03
U882C	8	Capsid	S	VP4	0.02
G2782A	7	Antigenic	NS	VP1	0.01
G491A	7	Noncoding	Noncoding	5' UTR	0.14
G619U	7	Noncoding	Noncoding	5' UTR	0
C2609U	6	Capsid	NS	VP1	0.01
C2783A	5	Antigenic	NS	VP1	0
U4374C	5	2C	S	2C	0.55
A3490G	4	2A	NS	2A	0.02
C2291U	4	Capsid	NS	VP3	0.01
C2580U	4	Capsid	S	VP1	0.1
G1282A	4	Capsid	NS	VP2	0
U1641C	4	Capsid	S	VP2	0.68
U5811A	4	3C	S	3C	0.14
U6693A	4	3D	S	3D	0.31

<sup>a</sup>Mutation presented as base in OPV2, position in OPV2 reference genome, and base in samples.

<sup>b</sup>Number of individuals with each mutation present at a frequency of 5% or greater. The total number of individuals analyzed is 83.

<sup>c</sup>Nonsynonymous (NS) or synonymous (S) mutations relative to the OPV2 reference genome.

<sup>d</sup>Fraction of cVDPV genomes with each mutation.

We applied two models to quantify the effective genetic bottleneck at transmission (Sobel Leonard et al., 2017; McCrone et al., 2018). We use the term “effective” bottleneck to clarify that we are capturing mutations that transmit and persist in the population of the recipient host. The presence-absence model asks whether a polymorphism in the donor is present or absent in the recipient, assuming perfect detection. Here, transmission is modeled as a random sampling process in which the probability of transmission is a function of the mutation frequency in the donor and the size of the transmission bottleneck. We used maximum likelihood optimization to find the bottleneck size distribution that best fit the data, assuming that the bottlenecks follow a zero-truncated Poisson distribution. Under the presence-absence model, the mean bottleneck size was 1.98 ( $\lambda = 1.57$ , 95% confidence interval 0.43–3.63), indicating that most bottlenecks are 2% and that 95% of bottlenecks are less than 4 (Figure 5B). We also applied a beta-binomial model, which incorporates the sensitivity of detecting variants in the recipient and allows for time-dependent stochastic loss of variants. The beta-binomial model yielded a mean bottleneck size of 2.11 ( $\lambda = 1.74$ ). The model fit was not significantly better than the presence-absence model (AIC 42.1 for presence-absence versus 39.7 for beta binomial), indicating that the loss of sensitivity might not be an important factor in these models.

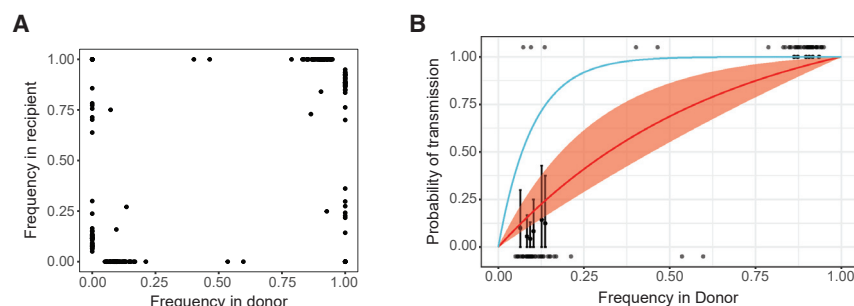
We also estimated bottlenecks for each pair individually with both models (Table S5). The two models produced the same estimates for each pair, although the beta-binomial model resulted in slightly larger confidence intervals. These results suggest that few genetically distinct OPV2 genomes transmit and persist in new hosts.

These models assume that minor variants are transmitted independently; however, variants can potentially be linked within hosts, which could result in an inflated bottleneck estimate. Because we used amplicon sequencing with short reads, we were unable to evaluate and exclude mutation linkage across the entire genome. However, we were able to evaluate linkage among pairs of intrahost variants that were near enough to be spanned by individual reads. These variant pairs were largely independent (Figure S5). Given these results and the fact that our bottleneck estimate was already low, the assumption of independent variant transmission is likely reasonable in this context.

### A Tight Transmission Bottleneck Limits the Spread of Gatekeeper Mutations

We sought to investigate whether a narrow transmission bottleneck would impact the transmission of mutations that are positively selected within vaccine recipients. Based on the estimated bottleneck size, we calculated the probability of transmission of





**Figure 5. Shared Viral Diversity across Transmission Pairs and Transmission Bottleneck**

(A) iSNV for four pairs of mOPV2 recipients and their household contacts. Each iSNV is plotted as a point with its frequency in the recipient (y axis) versus its frequency in the donor (x axis).

(B) Presence-absence bottleneck model fit compared with data. Frequency of donor iSNV on the x axis and probability of transmission on the y axis. Transmitted iSNV are shown along the top of the plot and non-transmitted iSNV are shown along the bottom. The red line shows the probability of transmission as a function of donor frequency given the mean bottleneck estimate, with a 95% confidence interval shown by the shaded area. The blue line shows the probability of transmission given a bottleneck size of 10 unique genomes. The black points on the graph represent the probability of transmission from the measured iSNV using a sliding window of 3% width and a step size of 1.5%.

each of the three gatekeeper mutations as a function of time, using the median frequency from mOPV2 recipients (Figure 6A). We then calculated the fraction of samples from transmission recipients that had each mutation present, regardless of whether we had sequence data from a donor population (Figure 6B). The fraction of transmission samples with each mutation is consistent with the calculated probability of transmission over time given the size of the bottleneck. We identified A481G in most transmission samples, consistent with its rapid fixation. However, few samples contained U2909C and no samples contained U398C. The majority of transmission events occurred within the first 2 weeks after the vaccination campaign, prior to when most vaccine recipients acquired U2909C and U398C (Taniuchi et al., 2017). These data suggest that the three gatekeeper mutations were not preferentially transmitted; instead, they suggest that mutations must rise to an appreciable frequency early enough within a donor population to be frequently transmitted through a narrow bottleneck.

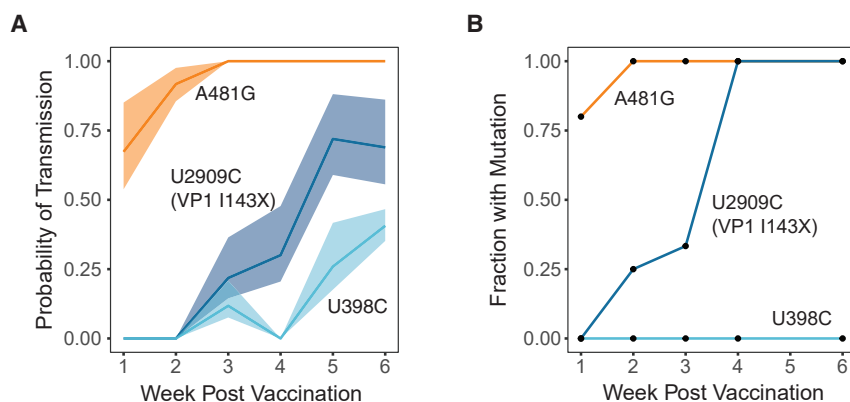
## DISCUSSION

We used whole genome deep sequencing to define the within-host evolutionary dynamics of OPV2 in a clinical trial in Matlab, Bangladesh (Taniuchi et al., 2017). The trial enabled analyses of longitudinal samples from a defined and synchronized point of mOPV2 vaccination and household transmission in a community with a high enteric pathogen burden and vaccine coverage. These results provide a rare window into the evolutionary dynamics that occur in the first weeks following vaccination. Similar to other RNA viruses, we identified strong purifying selection across the poliovirus genome within hosts (McCrone et al., 2018). However, in stark contrast to other viruses, we found evidence for strong within-host positive selection at multiple sites. Although high population immunity in Bangladesh limited the number of transmission samples available from the trial, we were able to quantify the transmission of key reversion mutations and estimate a tight bottleneck in this setting. Our findings enhance our knowledge on the within-host and transmission dynamics of polioviruses in relation to the development of cVDPV.

We found that positive selection is remarkably strong within vaccine recipients, with a magnitude that is seldom found in the within-host evolution of acute RNA viruses. We and others have rarely identified strong selection for mutations at the within-host level, even for mutations that should have beneficial

effects (Debbink et al., 2017; Dinis et al., 2016). The within-host evolution of several arboviruses is characterized by purifying selection and a large effect of stochastic genetic drift (Lequime et al., 2016; Parameswaran et al., 2012). In household cohort studies of influenza virus infection, we have found little evidence for positive selection within the span of a single infection (McCrone et al., 2018; Valesano et al., 2020), and iSNV are rarely observed in more than one individual. In contrast, in this cohort we identified 24 mutations that were identified in  $\geq 4$  individuals. Although comparisons to other viruses are complicated by differences in duration of infection, genome structure, and other factors, the extent of parallel evolution in OPV at this scale is remarkable. Whereas wild polioviruses and other endemic RNA viruses may already exist near local fitness peaks, OPV is significantly attenuated and is under intense pressure to climb the fitness landscape by accessing available high-impact mutations (Stern et al., 2017). OPV is also unique in that each population starts from the same founder genetic sequence, making parallel trajectories more likely to occur (Gutierrez et al., 2019).

Outside of the three gatekeeper mutations, we found that there are multiple additional sites under selection early in OPV2 evolution that reflect re-adaptation to the human host. There are several potential reasons why our study revealed mutations that have not been previously identified in cell culture or phylogenetic studies. While cell culture and animal models can be a helpful proxy for inferring selective pressures (Geoghegan and Holmes, 2018), they do not always capture the direction and magnitude of evolutionary forces in natural hosts. Similarly, phylogenetic studies have yielded important insights into the evolution and epidemiology of cVDPV but may not be able to infer the selective advantage of mutations with weaker effects due to limitations in sampling or statistical power. Whereas previous phylogenetic work on cVDPV2 found only limited evidence for positive selection at VP3-77 and VP1-222, they appear to be strongly selected within hosts (Shaw et al., 2018; Stern et al., 2017). Finally, it is important to recognize that phylogenetic studies may differ in the fitness effects they reveal due to differences in time and scale. Our results primarily reflect within-host selection observed through parallel evolution among vaccine recipients sampled longitudinally whereas previous phylogenetic analyses were based on shared variation among surveillance samples collected from many people over time and linked by sustained transmission. While we found that 16 of the 19 positively selected, non-gatekeeper variants recur in cVDPV



**Figure 6. Impact of a Tight Bottleneck on Transmission of Gatekeeper Mutations**

(A) The probability of transmission of each gatekeeper mutation calculated from the median frequency over time in the mOPV2 recipients given the estimated bottleneck. The shaded areas represent 95% confidence intervals based on the model fit. (B) The fraction of each gatekeeper mutation present above a frequency of 5% in samples from household contacts as a function of time since the vaccination campaign.

lineages, they do not routinely fix. Some mutations might be repeatedly selected within hosts and prove to be detrimental between them.

The underlying selective pressures and consequences of these mutations are unclear but their locations suggest functional significance. Mutations in the IRES have been shown to affect protein translation and replicative capacity (Avanzino et al., 2018). Mutations in the capsid enable adaptation to replication at physiologic temperatures and could provide increased structural stability or modulate receptor binding and viral entry (Macadam et al., 1991; Robinson et al., 2014). Although some of the capsid sites identified here are recognized by neutralizing antibodies (Patel et al., 1993; Shaw et al., 2018), there is little evidence that these mutations lead to antigenic escape, as they do in influenza virus or HIV (Karlsson Hedestam et al., 2008). Even highly diverged cVDPV strains with significant antigenic evolution are still neutralized by serum from vaccinated individuals (Shaw et al., 2018), and vaccination with OPV is used to control outbreaks of cVDPV (Kew and Pallansch, 2018). Rather than antigenic escape, we suggest that these mutations lead to improved within-host replication and, therefore, greater shedding and transmission. Epidemiologic data suggest that at some unknown point in cVDPV evolution, OPV achieves a level of transmissibility that is similar to that of wild polioviruses (Famulare et al., 2018; Jenkins et al., 2010; Duintjer Tebbens et al., 2013). Selection for phenotypes related to this increase in transmissibility, like enteric replication and shedding, are likely the earliest pressures the virus faces (Bull et al., 2018). Not all capsid antigenic sites may be involved in this process. There can be frequent amino acid substitution at many antigenic sites in the OPV2 capsid, often reverting back to previous replacements, suggesting that some sites are more tolerant to mutation and evolve more by genetic drift than selection (Shaw et al., 2018). However, our results indicate that a subset of these capsid sites experience positive selection and likely have functional effects related to improved replication and transmission within the human host.

Our identification of specific sites under positive selection has implications for genetic surveillance of VDPV. In VDPV isolates, the time since vaccine administration is estimated by molecular clock methods on VP1 sequence data (Jorba et al., 2008). For OPV2, the threshold for calling a strain a VDPV—as opposed to OPV-like—is 0.6% divergence, or  $\geq 6$  nucleotide substitutions

(Wassilak et al., 2011). Prior work has integrated the fixation rates of gatekeeper mutations into molecular clock models to refine estimates of the time between VDPV detection and initial vaccination (Famulare et al., 2016). By accounting for these rapidly selected mutations, the authors inferred that type 2 VDPVs are younger than estimated based on neutral evolution alone. Here, we used our longitudinal data to determine the fine-scale dynamics of these three gatekeeper mutations and to characterize the variability that can manifest at the individual scale. Fixation-rate estimates that are grounded in direct measurements are relevant for modeling efforts that rely on these parameters. In addition, we suggest that molecular clock models of VDPV might benefit from incorporation of the sites identified in this work in two ways: first, the rates of selection at these sites could be integrated into existing models of time since vaccine dosing for VDPVs; second, these sites under putative selection could be excluded from neutral molecular clocks for VDPV divergence time estimation.

A surprising and important finding is that a tight bottleneck (1–4 distinct genomes) limited the transmission of within-host variants to new hosts. It is certainly possible that loss of variants in the recipient population could result in an under-estimation of the bottleneck. In this study, we were limited by the weeklong interval between sample collection, which meant that transmission could have occurred several days prior to sample collection from household contacts. We also used a conservative frequency threshold of 5%, which may miss transmission of variants that remain at low frequencies across both hosts. However, it is unlikely that variants below 5% would be consistently transmitted while variants from 5%–50% are not. Furthermore, the results of the beta-binomial model suggest that imperfect detection and stochastic loss did not have a large influence on the bottleneck estimate.

In support of the finding of a small bottleneck in this transmission setting, we have no evidence for a between-host advantage of the gatekeeper mutations despite strong evidence of within-host selection. If genomes with gatekeeper mutations were transmitted preferentially, we would have expected to see them present in more household contacts and at higher frequencies than observed here. The sample size in this study limits our ability to detect small effects, but the data generally do not support substantial preferential transmission in this study population. A narrow, non-selective bottleneck can explain the

pattern of transmission to household contacts. In this scenario, mutations selected within vaccine recipients must rise above a threshold frequency prior to transmission in order to transit a narrow bottleneck. This would limit spread of minor iSNV that have not been selected quickly enough before finding a new host.

Of course, the gatekeeper mutations eventually become fixed in nearly all vaccine-derived lineages and can arise *de novo* in each subsequent host (Famulare et al., 2016; Stern et al., 2017). Population immunity and differences in fecal-oral exposure between the Matlab, which has never experienced a cVDPV outbreak, and other settings where cVDPV outbreaks are more common may lead to important selective differences not observed here. In contrast to the lack of between-host selective effects in this study, late transmission events at the end of the duration of shedding when variant fractions are high, may have a larger impact on the spread of positively selected mutations. Transmission, both later on in infections when viral load is low and to more distant community contacts, was uncommon in this highly immune population (Taniuchi et al., 2017). Furthermore, bottlenecks may be larger in populations with lower background immunity and higher fecal-oral exposure where naturally acquired doses are likely higher (Famulare et al., 2018), which would allow positive selection among minor variants to act.

It is likely that newly developed live-attenuated polio vaccines will face the same underlying selection pressures as mOPV2. Our results suggest that once a beneficial mutation occurs on a highly attenuated OPV background, there is strong selection to drive it to fixation. Strategies that decrease the fitness benefit of any single mutation, like modifications to IRES domain V and codon deoptimization, may be effective at sufficiently prolonging the time to reversion (Konopka-Anstadt et al., 2020; Yeh et al., 2020). However, modest decreases in mutation rate by introduction of high-fidelity RNA-dependent RNA polymerase modifications might have a lesser impact, as the mutation rate is still orders of magnitude higher than in other organisms (Sanjuán et al., 2010). In the setting of a virus starting from low fitness with a high mutation rate, whether a mutation achieves fixation or not may be more dependent on the extent of the fitness benefit rather than the waiting time for *de novo* generation of the mutation. This effect is illustrated by one next-generation OPV2 (nOPV2) design, which prevents A481G by modification of IRES domain V. In individuals receiving this nOPV2, VP1-143 and U398C still readily revert despite a high-fidelity 3D<sup>Pol</sup> (Yeh et al., 2020). High-fidelity polymerase modifications themselves may not be stable, as seen by the reversion and compensation of a type 1 poliovirus fidelity mutant, due to a fitness defect in cell culture (Fitzsimmons et al., 2018). Our results also suggest that there are mutations other than the three gatekeepers that increase fitness and contribute to reversion. Monitoring the genetic changes of new vaccine designs in sufficiently large cohorts will be important for evaluation of the genetic stability at these additional sites.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

### ● KEY RESOURCES TABLE

### ● RESOURCE AVAILABILITY

- Lead Contact
- Materials Availability
- Data and Code Availability

### ● EXPERIMENTAL MODEL AND SUBJECT DETAILS

- Clinical Trial Information and Ethics

### ● METHOD DETAILS

- Sample Collection and Viral Load Quantification
- Primer Design
- Amplification and Sequencing
- Benchmarking of Variant Identification
- Processing Sequence Data
- Identification of Within-Host Variants
- Permutation Test for Parallel Mutations
- Estimation of the Transmission Bottleneck

### ● QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.chom.2020.10.011>.

## ACKNOWLEDGMENTS

We thank the participants and field workers in the original vaccine trial for their many contributions. We acknowledge the following individuals for their roles in implementation of the vaccine trial and sharing of samples: Dr. Mohammed Yunus, M.B.B.S. (Division of Infectious Diseases, icddr,b, Dhaka, Bangladesh), and Dr. William A. Petri, M.D., Ph.D (Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA). We thank Suzanne Stroup for technical support. We thank Andrew Macadam, PhD and Cara Burns, PhD, for providing the OPV1 and OPV2 plasmids, respectively. We thank J.T. McCrone for helpful discussion on data analysis. This work was funded by a grant from the Bill and Melinda Gates Foundation (to M.F. and M.T.) and a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Diseases Award (to A.S.L.).

## AUTHOR CONTRIBUTIONS

M.T., M.F., and A.S.L. conceptualized the project. M.T., M.F., K.Z., and R.H. performed the vaccine trial described in prior work and provided the samples. A.L.V. performed the primer design and validation. A.L.V., W.J.F., M.O.I., and T.A. generated the sequence data. A.L.V. performed the sequence data analysis. A.S.L., W.W., and M.F. contributed to data analysis and data interpretation. A.L.V., M.T., M.F., and A.S.L. wrote and edited the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 2, 2020

Revised: October 12, 2020

Accepted: October 26, 2020

Published: November 18, 2020

## REFERENCES

- Acevedo, A., Brodsky, L., and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505, 686–690.
- Alam, N., Ali, T., Razzaque, A., Rahman, M., Zahirul Haq, M., Saha, S.K., Ahmed, A., Sarder, A.M., Moinuddin Haider, M., Yunus, M., et al. (2017). Health and demographic surveillance system (HDSS) in Matlab, Bangladesh. *Int. J. Epidemiol.* 46, 809–816.

- Avanzino, B.C., Jue, H., Miller, C.M., Cheung, E., Fuchs, G., and Fraser, C.S. (2018). Molecular mechanism of poliovirus Sabin vaccine strain attenuation. *J. Biol. Chem.* 293, 15471–15482.
- Blake, I.M., Pons-Salort, M., Molodecky, N.A., Diop, O.M., Chenoweth, P., Bandyopadhyay, A.S., Zaffran, M., Sutter, R.W., and Grassly, N.C. (2018). Type 2 poliovirus detection after global withdrawal of trivalent oral vaccine. *N. Engl. J. Med.* 379, 834–845.
- Boot, H.J., Sonsma, J., van Nunen, F., Abbink, F., Kimman, T.G., and Buisman, A.M. (2007). Determinants of monovalent oral polio vaccine mutagenesis in vaccinated elderly people. *Vaccine* 25, 4706–4714.
- Bull, J.J., Smithson, M.W., and Nuismer, S.L. (2018). Transmissible viral vaccines. *Trends Microbiol.* 26, 6–15.
- Burns, C.C., Diop, O.M., Sutter, R.W., and Kew, O.M. (2014). Vaccine-derived polioviruses. *J. Infect. Dis.* 210 (Supplement 1), S283–S293.
- Burns, C.C., Shaw, J., Jorba, J., Bukbuk, D., Adu, F., Gumedé, N., Pate, M.A., Abanida, E.A., Gasasira, A., Iber, J., et al. (2013). Multiple independent emergences of Type 2 vaccine-derived polioviruses during a large outbreak in Northern Nigeria. *J. Virol.* 87, 4907–4922.
- Debbink, K., McCrone, J.T., Petrie, J.G., Truscon, R., Johnson, E., Mantlo, E.K., Monto, A.S., and Luring, A.S. (2017). Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLoS Pathog.* 13, e1006194.
- Dedepsidis, E., Karakasiliotis, I., Paximadi, E., Kyriakopoulou, Z., Komiotis, D., and Markoulatos, P. (2006). Detection of unusual mutation within the VP1 region of different re-isolates of poliovirus Sabin vaccine. *Virus Genes* 33, 183–191.
- Dinis, J.M., Florek, N.W., Fatola, O.O., Moncla, L.H., Mutschler, J.P., Charlier, O.K., Meece, J.K., Belongia, E.A., and Friedrich, T.C. (2016). Deep sequencing reveals potential antigenic variants at low frequencies in influenza A virus-infected humans. *J. Virol.* 90, 3355–3365.
- Dolan, P.T., Whitfield, Z.J., and Andino, R. (2018a). Mechanisms and concepts in RNA virus population dynamics and evolution. *Annu. Rev. Virol.* 5, 69–92.
- Dolan, P.T., Whitfield, Z.J., and Andino, R. (2018b). Mapping the evolutionary potential of RNA viruses. *Cell Host Microbe* 23, 435–446.
- Duintjer Tebbens, R.J.D., Pallansch, M.A., Kim, J.H., Burns, C.C., Kew, O.M., Oberste, M.S., Diop, O.M., Wassilak, S.G.F., Cochi, S.L., and Thompson, K.M. (2013). Oral poliovirus vaccine evolution and insights relevant to modeling the risks of circulating vaccine-derived polioviruses (cVDPVs). *Risk Anal.* 33, 680–702.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Elena, S.F., Sanjuán, R., Bordería, A.V., and Turner, P.E. (2001). Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. *Infect. Genet. Evol.* 1, 41–48.
- Famulare, M., Chang, S., Iber, J., Zhao, K., Adeniji, J.A., Bukbuk, D., Baba, M., Behrend, M., Burns, C.C., and Oberste, M.S. (2016). Sabin vaccine reversion in the field: a comprehensive analysis of Sabin-like poliovirus isolates in Nigeria. *J. Virol.* 90, 317–331.
- Famulare, M., Selinger, C., McCarthy, K.A., Eckhoff, P.A., and Chabot-Couture, G. (2018). Assessing the stability of polio eradication after the withdrawal of oral polio vaccine. *PLoS Biol.* 16, e2002468.
- Fitzsimmons, W.J., Woods, R.J., McCrone, J.T., Woodman, A., Arnold, J.J., Yennawar, M., Evans, R., Cameron, C.E., and Luring, A.S. (2018). A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol.* 16, e2006459.
- Geoghegan, J.L., and Holmes, E.C. (2018). Evolutionary virology at 40. *Genetics* 210, 1151–1162.
- Gerstung, M., Beisel, C., Rechsteiner, M., Wild, P., Schraml, P., Moch, H., and Beerewinkel, N. (2012). Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat. Commun.* 3, 811.
- Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 20, 8.
- Gutierrez, B., Escalera-Zamudio, M., and Pybus, O.G. (2019). Parallel molecular evolution and adaptation in viruses. *Curr. Opin. Virol.* 34, 90–96.
- Holubar, M., Sahoo, M.K., Huang, C., Mohamed-Hadley, A., Liu, Y., Waggoner, J.J., Troy, S.B., García-García, L., Ferreyra-Reyes, L., Maldonado, Y., et al. (2019). Deep sequencing prompts the modification of a real-time RT-PCR for the serotype-specific detection of polioviruses. *J. Virol. Methods* 264, 38–43.
- Jenkins, H.E., Aylward, R.B., Gasasira, A., Donnelly, C.A., Mwanza, M., Corander, J., Garnier, S., Chauvin, C., Abanida, E., Pate, M.A., et al. (2010). Implications of a circulating vaccine-derived poliovirus in Nigeria. *N. Engl. J. Med.* 362, 2360–2369.
- Jorba, J., Campagnoli, R., De, L., and Kew, O. (2008). Calibration of multiple poliovirus molecular clocks covering an extended evolutionary range. *J. Virol.* 82, 4429–4440.
- Karlsson Hedestam, G.B.K., Fouchier, R.A.M., Phogat, S., Burton, D.R., Sodroski, J., and Wyatt, R.T. (2008). The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat. Rev. Microbiol.* 6, 143–155.
- Kew, O., Morris-Glasgow, V., Landaverde, M., Burns, C., Shaw, J., Garib, Z., André, J., Blackman, E., Freeman, C.J., Jorba, J., et al. (2002). Outbreak of poliomyelitis in Hispaniola associated with circulating Type 1 vaccine-derived poliovirus. *Science* 296, 356–359.
- Kew, O., and Pallansch, M. (2018). Breaking the last chains of poliovirus transmission: progress and challenges in global polio eradication. *Annu. Rev. Virol.* 5, 427–451.
- Kew, O.M., Sutter, R.W., de Gourville, E.M., Dowdle, W.R., and Pallansch, M.A. (2005). Vaccine-derived polioviruses and the endgame strategy for global polio eradication. *Annu. Rev. Microbiol.* 59, 587–635.
- Kilpatrick, D.R., Iber, J.C., Chen, Q., Ching, K., Yang, S.J., De, L., Mandelbaum, M.D., Emery, B., Campagnoli, R., Burns, C.C., et al. (2011). Poliovirus serotype-specific VP1 sequencing primers. *J. Virol. Methods* 174, 128–130.
- Konopka-Anstadt, J.L., Campagnoli, R., Vincent, A., Shaw, J., Wei, L., Wynn, N.T., Smithee, S.E., Bujaki, E., Te Yeh, M., Laassri, M., et al. (2020). Development of a new oral poliovirus vaccine for the eradication end game using codon deoptimization. *NPJ Vaccines* 5, 26.
- Kryazhimskiy, S., and Plotkin, J.B. (2008). The population genetics of dN/dS. *PLoS Genet.* 4, e1000304.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Luring, A.S. (2020). Within-host viral diversity: a window into viral evolution. *Annu. Rev. Virol.* 7, 63–81.
- Lequime, S., Fontaine, A., Ar Gouilh, M., Moltini-Conclois, I., and Lambrechts, L. (2016). Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. *PLoS Genet.* 12, e1006111.
- Lévesque, N., and Semler, B.L. (2015). A 21st century perspective of poliovirus replication. *PLoS Pathog.* 11, e1004825.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Macadam, A.J., Pollard, S.R., Ferguson, G., Dunn, G., Skuce, R., Almond, J.W., and Minor, P.D. (1991). The 5′ noncoding region of the type 2 poliovirus vaccine strain contains determinants of attenuation and temperature sensitivity. *Virology* 181, 451–458.
- Macadam, A.J., Pollard, S.R., Ferguson, G., Skuce, R., Wood, D., Almond, J.W., and Minor, P.D. (1993). Genetic basis of attenuation of the Sabin Type 2 vaccine strain of poliovirus in Primates. *Virology* 192, 18–26.
- Macklin, G.R., O'Reilly, K.M., Grassly, N.C., Edmunds, W.J., Mach, O., Krishnan, R.S.G., Voorman, A., Vertefeuille, J.F., Abdelwahab, J., Gumedé, N., et al. (2020). Evolving epidemiology of poliovirus serotype 2 following withdrawal of the type 2 oral poliovirus vaccine. *Science* 368, 401–405.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12.



- McCrone, J.T., and Lauring, A.S. (2016). Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. *J. Virol.* **90**, 6884–6895.
- McCrone, J.T., Woods, R.J., Martin, E.T., Malosh, R.E., Monto, A.S., and Lauring, A.S. (2018). Stochastic processes constrain the within and between host evolution of influenza virus. *eLife* **7**, e35962.
- Montmayeur, A.M., Ng, T.F.F., Schmidt, A., Zhao, K., Magaña, L., Iber, J., Castro, C.J., Chen, Q., Henderson, E., Ramos, E., et al. (2017). High-throughput next-generation sequencing of polioviruses. *J. Clin. Microbiol.* **55**, 606–615.
- Muzychenko, A.R., Lipskaya, G.Yu., G.Yu., Maslova, S.V., Svitkin, Y.V., Pilipenko, E.V., Nottay, B.K., Kew, O.M., and Agol, V.I. (1991). Coupled mutations in the 5'-untranslated region of the Sabin poliovirus strains during in vivo passages: structural and functional implications. *Virus Res.* **21**, 111–122.
- Neverov, A., and Chumakov, K. (2010). Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines. *Proc. Natl. Acad. Sci. USA* **107**, 20063–20068.
- Parameswaran, P., Charlebois, P., Tellez, Y., Nunez, A., Ryan, E.M., Malboeuf, C.M., Levin, J.Z., Lennon, N.J., Balmaseda, A., Harris, E., et al. (2012). Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. *J. Virol.* **86**, 8546–8558.
- Patel, V., Ferguson, M., and Minor, P.D. (1993). Antigenic sites on Type 2 poliovirus. *Virology* **192**, 361–364.
- Pons-Salort, M., Molodecky, N.A., O'Reilly, K.M., Wadood, M.Z., Safdar, R.M., Etsano, A., Vaz, R.G., Jafari, H., Grassly, N.C., and Blake, I.M. (2016). Population immunity against Serotype-2 poliomyelitis leading up to the global withdrawal of the oral poliovirus vaccine: spatio-temporal modelling of surveillance data. *PLoS Med.* **13**, e1002140.
- Ren, R.B., Moss, E.G., and Racaniello, V.R. (1991). Identification of two determinants that attenuate vaccine-related type 2 poliovirus. *J. Virol.* **65**, 1377–1382.
- Robinson, C.M., Jesudhasan, P.R., and Pfeiffer, J.K. (2014). Bacterial lipopolysaccharide binding enhances virion stability and promotes environmental fitness of an enteric virus. *Cell Host Microbe* **15**, 36–46.
- Sahoo, M.K., Holubar, M., Huang, C., Mohamed-Hadley, A., Liu, Y., Waggoner, J.J., Troy, S.B., Garcia-Garcia, L., Ferreyra-Reyes, L., Maldonado, Y., et al. (2017). Detection of emerging vaccine-related polioviruses by deep sequencing. *J. Clin. Microbiol.* **55**, 2162–2171.
- van der Sanden, S., Pallansch, M.A., van de Kasstele, J., El-Sayed, N., Sutter, R.W., Koopmans, M., and van der Avoort, H. (2009). Shedding of vaccine viruses with increased antigenic and genetic divergence after vaccination of newborns with monovalent Type 1 oral poliovirus vaccine. *J. Virol.* **83**, 8693–8704.
- Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral mutation rates. *J. Virol.* **84**, 9733–9748.
- Sarcey, E., Serres, A., Tindy, F., Chareyre, A., Ng, S., Nicolas, M., Vetter, E., Bonnevey, T., Abachin, E., and Mallet, L. (2017). Quantifying low-frequency revertants in oral poliovirus vaccine using next generation sequencing. *J. Virol. Methods* **246**, 75–80.
- Shaw, J., Jorba, J., Zhao, K., Iber, J., Chen, Q., Adu, F., Adeniji, A., Bukbuk, D., Baba, M., Henderson, E., et al. (2018). Dynamics of evolution of poliovirus neutralizing antigenic sites and other capsid functional domains during a large and prolonged outbreak. *J. Virol.* **92**, e01949-17.
- Sobel Leonard, A.S., Weissman, D.B., Greenbaum, B., Ghedin, E., and Koelle, K. (2017). Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *J. Virol.* **91**, e00171-17.
- Stern, A., Yeh, M.T., Zinger, T., Smith, M., Wright, C., Ling, G., Nielsen, R., Macadam, A., and Andino, R. (2017). The evolutionary pathway to virulence of an RNA virus. *Cell* **169**, 35–46.e19.
- Taniuchi, M., Famulare, M., Zaman, K., Uddin, M.J., Upfill-Brown, A.M., Ahmed, T., Saha, P., Haque, R., Bandyopadhyay, A.S., Modlin, J.F., et al. (2017). Community transmission of type 2 poliovirus after cessation of trivalent oral polio vaccine in Bangladesh: an open-label cluster-randomised trial and modelling study. *Lancet Infect. Dis.* **17**, 1069–1079.
- Valesano, A.L., Fitzsimmons, W.J., McCrone, J.T., Petrie, J.G., Monto, A.S., Martin, E.T., and Lauring, A.S. (2020). Influenza B viruses exhibit lower within-host diversity than influenza A viruses in human hosts. *J. Virol.* **94**, e01710–e01719.
- Wassilak, S., Pate, M.A., Wannemuehler, K., Jenks, J., Burns, C., Chenoweth, P., Abanida, E.A., Adu, F., Baba, M., Gasasira, A., et al. (2011). Outbreak of Type 2 vaccine-derived poliovirus in Nigeria: emergence and widespread circulation in an underimmunized population. *J. Infect. Dis.* **203**, 898–909.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.
- Yang, Z., Wong, W.S.W., and Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118.
- Yeh, M.T., Bujaki, E., Dolan, P.T., Smith, M., Wahid, R., Konz, J., Weiner, A.J., Bandyopadhyay, A.S., Van Damme, P., De Coster, I., et al. (2020). Engineering the live-attenuated polio vaccine to prevent reversion to virulence. *Cell Host Microbe* **27**, 736–751.e8.
- Yoon, H., and Leitner, T. (2015). PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. *Bioinformatics* **31**, 1472–1474.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and Virus Strains</b>		
Poliovirus type 1 Mahoney strain	Clone was gift of Eckard Wimmer (Stony Brook University, Stony Brook, NY)	N/A
Oral polio vaccine, Sabin strain serotype 1	Clone was gift of Dr. Andrew Macadam (National Institute for Biological Standards and Control, London, UK)	N/A
<b>Biological Samples</b>		
Total nucleic acids from stool from clinical trial in Matlab, Bangladesh	This paper; fully described in <a href="#">Taniuchi et al., 2017</a>	N/A
Total nucleic acids from stool from healthy, de-identified donors, used in sequencing validation experiment	Gift of Pat Schloss (University of Michigan, Ann Arbor, MI)	N/A
<b>Critical Commercial Assays</b>		
QIAamp Viral RNA Mini Kit	Qiagen	Cat#52904
SuperScript III First-Strand Synthesis System	Thermo Fisher	Cat#18080051
Phusion High-Fidelity DNA Polymerase	NEB	Cat#M0530L
Agencourt AMPure XP magnetic beads	Beckman Coulter	Cat#A63881
Quant-iT™ PicoGreen™ dsDNA Assay Kit	Thermo Fisher	Cat#P7589
Nextera DNA Flex Library Prep Kit	Illumina	Cat#20018705
Nextera DNA CD Indexes (96 Indexes, 96 Samples)	Illumina	Cat#20018708
<b>Deposited Data</b>		
Raw sequence read files (fastq)	NCBI Sequence Read Archive	BioProject accession PRJNA637613
<b>Oligonucleotides</b>		
Sequences of primers used for poliovirus genome amplification	<a href="#">Table S1</a> of this paper	N/A
<b>Recombinant DNA</b>		
Sabin strain serotype 2 clone in pUC19 plasmid vector, used as sequencing control	Gift from Dr. Cara Burns (Center for Disease Control and Prevention, Atlanta, GA)	N/A
<b>Software and Algorithms</b>		
cutadapt	<a href="#">Martin, 2011</a>	<a href="https://cutadapt.readthedocs.io/en/stable/">https://cutadapt.readthedocs.io/en/stable/</a>
Bowtie2	<a href="#">Langmead and Salzberg, 2012</a>	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
samtools	<a href="#">Li et al., 2009</a>	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
MUSCLE	<a href="#">Edgar, 2004</a>	<a href="https://www.drive5.com/muscle/manual/">https://www.drive5.com/muscle/manual/</a>
PAML	<a href="#">Yang, 2007</a>	<a href="http://abacus.gene.ucl.ac.uk/software/paml.html">http://abacus.gene.ucl.ac.uk/software/paml.html</a>
Analysis code	This paper (GitHub)	<a href="https://github.com/lauringlab/Poliovirus_Intrahost">https://github.com/lauringlab/Poliovirus_Intrahost</a>

### RESOURCE AVAILABILITY

#### Lead Contact

Requests for further information should be directed to the lead contact, Adam Lauring ([alauring@med.umich.edu](mailto:alauring@med.umich.edu)).

## Materials Availability

This study did not produce new unique reagents or materials.

## Data and Code Availability

The raw sequence reads for the MATLAB samples and the benchmarking experiment are available on the NCBI Sequence Read Archive in BioProject PRJNA637613. Reads aligning to the human genome were filtered out by the SRA. The code for the primary analysis of within-host variants is publicly available at [https://github.com/lauringlab/variant\\_pipeline](https://github.com/lauringlab/variant_pipeline). The rest of the code for data analysis and generation of the figures was written in R version 3.5.0 and python2.7 and is publicly available on GitHub at [https://github.com/lauringlab/Poliovirus\\_Intrahost](https://github.com/lauringlab/Poliovirus_Intrahost).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Clinical Trial Information and Ethics

The clinical trial, including all aspects of sample collection and viral load measurements, is described in full in a prior publication (Taniuchi et al., 2017). The study was done according to the guidelines of the Declaration of Helsinki. The protocol for the clinical trial was approved by the Research Review Committee (RRC) and Ethical Review Committee (ERC) of the icddr,b and the Institutional Review Board of the University of Virginia. It is registered at [ClinicalTrials.gov](https://clinicaltrials.gov), number NCT02477046.

## METHOD DETAILS

### Sample Collection and Viral Load Quantification

Stool sample collection, nucleic acid extraction, and viral load measurements were performed and described in a prior publication (Taniuchi et al., 2017). Briefly, stool samples were collected, placed at 4°C, and delivered to the icddr,b laboratory in MATLAB within 6 h. Samples were then aliquoted and stored at -80°C until shipment on dry ice to the icddr,b laboratories in Dhaka. Total nucleic acid (TNA) from approximately 200 grams of stool was extracted with the QIAamp Fast DNA Stool mini kit and OPV was detected and quantified by RT-qPCR with serotype specific primers. TNA samples were shipped on dry ice to the University of Virginia and stored at -80 °C until processing for sequencing.

### Primer Design

We designed primers to amplify all three serotypes of OPV in overlapping amplicons covering the poliovirus genome. We used PrimerDesign-M (Yoon and Leitner, 2015) to determine sites of conservation across the three poliovirus types and identify potential primer sequences, allowing for ambiguous bases. We selected primers such that the four segments overlapped by at least 500 bp and manually curated each primer to have similar melting temperatures. We empirically tested various primer candidates for amplification on type 1 and type 2 poliovirus RNA templates. The primers used for genome amplification are listed in Table S1.

### Amplification and Sequencing

We amplified viral cDNA in four amplicons using a two-step RT-PCR protocol. We performed reverse transcription using the SuperScript III First-Strand Synthesis System (Thermo Fisher). Each reaction contained 1.13  $\mu$ L of 50 ng/ $\mu$ L random hexamers, 0.37  $\mu$ L oligo-dT, 1.5  $\mu$ L 10 mM dNTP mix, 12  $\mu$ L of template total nucleic acid from stool, 3  $\mu$ L of 10x RT Buffer, 6  $\mu$ L 25 mM MgCl<sub>2</sub>, 3  $\mu$ L 0.1M DTT, 1.5  $\mu$ L RNase Out, and 1.5  $\mu$ L of SuperScript III RT enzyme. The mixture of template, primer, and dNTPs was heated at 75°C for 15 min to denature RNA secondary structure and placed directly on ice for > 1 minute. The enzyme and buffer mixture were then added on ice. The thermocycler protocol was: 25°C for 10 min, 50°C for 50 min, 85°C for 5 min, and hold at 4°C. The four overlapping segments were amplified by PCR with the primers listed in Table S1. The PCR reactions were as follows: 10  $\mu$ L 5x HF Buffer, 1  $\mu$ L 10 mM dNTP mix, 0.25  $\mu$ L 100  $\mu$ M forward primer, 0.25  $\mu$ L 100  $\mu$ M reverse primer, 33  $\mu$ L nuclease free water, 0.5  $\mu$ L of Phusion DNA Polymerase (NEB), and 5  $\mu$ L of template cDNA. The thermocycler protocol was: 98°C for 30 sec, 40 cycles of 98°C for 10 sec, 59.5°C for 30 sec, 72°C for 2 min, then 72°C for 2 min for final extension, and hold at 4°C. The four segments for each sample were pooled in equal volumes (18  $\mu$ L of each segment for a total pooled volume of 72  $\mu$ L). Pooled amplicons were purified with Agencourt AMPure XP magnetic beads, using 1.8X volume of beads (129.6  $\mu$ L of beads for 72  $\mu$ L pooled PCR product). The sample was eluted into 40  $\mu$ L of nuclease-free water. The purified PCR products were quantitated by Quant-iT PicoGreen dsDNA High Sensitivity Assay. A limited number of PCR products were spot-checked by gel electrophoresis. A plasmid control was prepared by applying the PCR protocol to a template of OPV2 in a plasmid. The sequence of the plasmid was determined by Sanger sequencing and was identical to the OPV2 GenBank reference (AY184220.1). One plasmid control was included in each pooled library, beginning at the library preparation stage, to account for sequencing errors and batch effects. Samples between 9 x 10<sup>5</sup> copies/gram and 4.5 x 10<sup>7</sup> copies/gram of OPV2 were amplified and sequenced in duplicate to improve the specificity of within-host variant identification. Libraries were prepared for Illumina sequencing with the Nextera DNA Flex Library Preparation kit according to the manufacturer's instructions, using Nextera DNA CD Indexes (96 samples). Eight pooled libraries were prepared in total and sequenced on an Illumina MiSeq (2x250 reads, v2 chemistry).

### Benchmarking of Variant Identification

To determine the sensitivity and specificity of variant identification, we sequenced mock populations of a mixture of two viruses using the protocol described above. The viruses used were wild-type Mahoney type 1 poliovirus and the type 1 OPV strain, which differ by 66 mutations within the amplified regions. The consensus sequences of each viral stock were confirmed by Sanger sequencing. Viral RNA was extracted from each stock with the QIAamp Viral RNA Mini Kit (Qiagen) and viral RNA were mixed in equal concentrations at 0%, 1%, 2%, 5%, 10%, and 100% WT in OPV1. Virus mixtures were diluted to genome copy concentrations of  $4.5 \times 10^4$  copies/ $\mu$ L,  $9 \times 10^3$  copies/ $\mu$ L,  $9 \times 10^2$  copies/ $\mu$ L, and  $9 \times 10^1$  copies/ $\mu$ L (copies/gram is related to copies/ $\mu$ L by a factor of  $10^3$ ). To simulate the complex mixture of nucleic acid present in our samples, we performed the dilutions of viral populations in total nucleic acid extracted from stool from deidentified human donors (a gift of Pat Schloss, University of Michigan). The mixtures were then amplified by the protocol described in the section above (*Amplification and sequencing*). A plasmid control was generated from the OPV1 plasmid clone in the same way as described in the section above (*Amplification and sequencing*). The pooled library was generated using the Nextera DNA Flex Library Preparation Kit and sequenced on an Illumina MiSeq (2x250 reads, v2 chemistry), including the OPV1 plasmid control to account for batch effects and errors. To more carefully estimate the sensitivity of variant identification at various coverage levels, mapped reads were randomly down sampled to approximately 1000x, 500x, and 200x coverage evenly across the genome. Then within-host variants were identified using an analytic pipeline previously used for influenza viruses (McCrone et al., 2018; McCrone and Lauring, 2016), which depends on a clonal plasmid control to account for batch effects and local errors in Illumina sequencing.

### Processing Sequence Data

Sequencing adapters were removed with cutadapt (Martin, 2011) and reads were aligned to all three OPV reference genomes (AY184220.1, AY184221.1, V01150.1) using bowtie2 (Langmead and Salzberg, 2012) with the *-very-sensitive* option. Duplicate reads were removed with Picard and samtools (Li et al., 2009). Consensus bases were identified at sites with 10x coverage or greater. For samples sequenced in duplicate, the replicate with the higher coverage at a given site was used to assign the consensus base. Each biological sample was assigned to a group based on depth and evenness of coverage across the OPV2 reference. Mean coverage was calculated across non-overlapping 50 bp bins across the genome region amplified by the four amplicon segments. Samples were considered variant-quality if they had an average coverage greater than 200x in every bin. Samples that had coverage of 10x or greater at every site were considered consensus-quality. We aligned the consensus sequences with the OPV2 reference with the MUSCLE algorithm (Edgar, 2004). For the dN/dS analysis, we used the PAML software version 4.8 (Yang, 2007). For gene-wise dN/dS analysis, we calculated a single value for omega across each gene with codeml using model M0. To identify sites with evidence of positive selection, we compared the likelihood of models M2 vs M1 with a chi-squared test and identified sites with omega greater than 1 with the Bayes empirical Bayes method.

### Identification of Within-Host Variants

We identified within-host variants in any 50 bp window with greater than 200x mean coverage, even if fewer than four segments were successfully amplified and sequenced. Within-host variants on the OPV2 genome were identified with the R package deepSNV (Gerstung et al., 2012), using the OPV2 plasmid control to account for sequencing errors and strand bias. Minor iSNV (< 50% frequency) in the cohort samples were filtered using the following criteria: deepSNV p value < 0.01, average mapping quality > 20, average Phred score > 35, and average read position in the middle 75% of the read (positions 31 and 219 for 250 bp pair reads). For samples sequenced in duplicate, we only used variants identified in both samples; we assigned frequency using the sample that had higher coverage at the site. We only identified iSNV present at a frequency of > 5%. Sites that were monomorphic after applying these filter criteria were assigned a frequency of 100%. The analytic pipeline was used to determine the position of each base in the coding sequence of the viral polyprotein and assign it as synonymous or non-synonymous relative to the sample consensus.

To obtain haplotype information specifically for VP1-143, codon frequency was identified by finding all reads that spanned the codon, filtering by MapQ > 20 and Phred score of each base > 20, counting the number of reads within each codon, and dividing by the number of reads that passed the quality filters. Samples with quality read depth less than 150 were excluded.

### Permutation Test for Parallel Mutations

We quantified the probability that mutations would arise in parallel across a given number of individuals by implementing a permutation test. We first assumed that all sites were equally likely to mutate. We found the number of mutations that occurred in 83 individuals with variant-quality samples relative to the OPV2 reference above a frequency of 5%. We used this distribution to draw sites randomly across the genome, accounting for the length of the region amplified in our assay and excluding primer binding sites. We then found the number of sites shared by a given number of individuals. We ran this permutation 1000 times and calculated the p value as the number of permutations with a number of shared sites equal to or greater than the observed data for a given group (e.g. mutations shared by two individuals, etc.). We simulated constraint on the mutability of genomic sites by restricting the fraction of sites available to mutate. We chose a fraction available of 60% to reflect the known distribution of fitness effects in poliovirus based on experimental data (Acevedo et al., 2014).

### Estimation of the Transmission Bottleneck

Models for estimating the transmission bottleneck were implemented as described in our prior work on influenza virus (McCrone et al., 2018). In the presence-absence model, we assessed whether donor iSNV are found in the recipient. We assumed perfect detection of transmitted iSNV and that the probability of transmission of donor iSNV are determined by the measured frequency at the time of sampling. We modeled the probability of transmission as a binomial sampling process, depending on the donor iSNV frequency and the bottleneck size ( $N_b$ ). We used maximum likelihood optimization to estimate the bottleneck size distribution, assuming bottlenecks across pairs follow a zero-truncated Poisson distribution. In the beta-binomial model, we relaxed the assumption of perfect detection of iSNV in the recipient by accounting for false-negative variant calls and stochastic loss below our detection threshold. We use our benchmarking data to supply the sensitivity of variant identification by frequency and titer, rounding down to the nearest titer threshold (e.g.  $4.5 \times 10^4$  copies/ $\mu$ L,  $9 \times 10^3$  copies/ $\mu$ L, etc.). We assume sensitivity in each range is the same as that of the titer threshold.

### QUANTIFICATION AND STATISTICAL ANALYSIS

We performed various statistical tests on the data, all of which are described in the Results and Method Details. Unless otherwise noted, statistical tests were performed in R version 3.5.0. We used a multiple linear model to measure the effects of time since vaccination and viral load on specimen iSNV richness shown in Figure 2B ( $n = 101$  specimens). We used a linear regression model to quantify the precision of iSNV frequency measurements across 11 variant-quality specimens sequenced in duplicate (Figure S2A). We applied a beta regression model (R package “betareg”) with a logit link function to the mutation frequency data shown in Figures 3A and S3B for three mutations. For samples with a frequency of 0 or 1, we adjusted their frequency by  $10^{-7}$  in order to apply the beta regression model. For the dN/dS analyses, we used all complete consensus genomes from mOPV2 vaccine recipients ( $n = 157$  genomes). For dN/dS calculations, we used PAML version 4.8 (Yang, 2007). For codon-specific dN/dS analyses, we used the Bayes empirical Bayes method (Yang et al., 2005). The permutation test for parallel mutations and the transmission bottleneck analyses are described in the Method Details.